Random Hyperplane LSH: Angular Probability and Entropy

Random hyperplane locality-sensitive hashing (LSH) generates binary codes by projecting an input vector **x** onto random directions **r** and recording only the sign:

$$b = \text{sign}(r \cdot x)$$

where each random vector **r** defines a hyperplane through the origin.

Angular collision probability

A fundamental result due to Charikar (2002) states that for two vectors **x** and **y** separated by angle θ,

$$P(b_x \neq b_y) = \theta/\pi$$

where θ is measured in radians.

Equivalently,

$$P(b_x = b_y) = 1 - \theta/\pi.$$

Geometrically, a random hyperplane separates **x** and **y** if its orientation falls within the angular wedge between them. Since the measure of the separating orientations is proportional to θ, the disagreement probability is exactly θ/π.

Consequently, for a k-bit hash code,

$$E[d_H(b_x, b_y)] = k\,\theta/\pi$$

where d_H denotes Hamming distance.

Thus Hamming distance between LSH codes provides an unbiased estimator of angular distance.

Information-theoretic interpretation

Each hyperplane asks a one-bit geometric question:

> On which side of this hyperplane does the input lie?

If the projection distribution is approximately symmetric,

$P(b = 0) \approx P(b = 1) \approx 0.5$

and the bit entropy approaches its maximum value:

$H(b) = -p \log_2(p) - (1-p) \log_2(1-p) \approx 1$ bit.

Random hyperplane LSH may therefore be viewed as a sequence of approximately maximally informative binary measurements of input geometry.

For two nearby inputs,

$\theta \approx 0$

implies

$P(bx \neq by) \approx 0,$

so the mutual information between their codes is high.

For orthogonal inputs,

$\theta = \pi/2$

implies

$P(bx \neq by) = 0.5,$

and the bits become maximally uncertain relative to one another.

Hence angular separation is directly converted into code entropy.

Relevance to Atlas LSH Neural Networks

In Atlas-style architectures, LSH bits act as conditional routing signals controlling parameter selection, memory access, or computational paths.

The network may therefore be interpreted as repeatedly performing:

1. approximately one-bit geometric measurements,
2. entropy reduction concerning the input location on the data manifold,
3. conditional computation based on the accumulated geometric information.

Under this view, Atlas networks resemble deep hierarchies of entropy-extracting geometric queries, where computation is routed according to progressively refined estimates of input position in representation space.

Reference

Moses Charikar, *Similarity Estimation Techniques from Rounding Algorithms*, Proceedings of the 34th Annual ACM Symposium on Theory of Computing (STOC), 2002.